

- 8 -

## REMARKS

In response to the above-identified Office action, Applicants have amended the specification; amended claims 1-10 and 35-36; and added new claims 42-48, as shown above and explained further herein below. Applicants hereby request further examination and allowance of claims 1-10, 35-36 and 42-48 in view of these amendments and the following remarks.

The Office has objected to the specification asserting that the abstract is incomplete and requires submission of a new abstract. In response, Applicants submit a new abstract. Support for the new abstract is found in the above-identified application at pages 4-5; pages 19-25; FIGS. 5-6; and claims 1-41 as originally filed. Accordingly, no new matter has been added. As such, the Office is respectfully requested to enter this new abstract and withdraw the objection.

The Office has rejected claims 1-10, 35 and 36 under 35 U.S.C. § 112, first paragraph, asserting that the language, “wherein the usage of a given tuple relates to a *relative frequency* with which the given tuple was accessed by the queries in the workload,” as recited in claims 1, 7, 10 and 35, was not described in the specification in such a way as to enable one skilled in the art to make or use the invention (emphasis added). In response, Applicants have amended claims 1-10, 35 and 36 as shown above. Support for these amendments is found in the above-identified application at pages 4-5; pages 19-25; FIGS. 5-6; and claims 1-41 as originally filed. Accordingly, no new matter has been added. Further, these amendments do not change the scope of the claims and make explicit what was implicit in the prior language to clarify the claims. It is believed these amendments alone overcome the outstanding rejections. Nevertheless, Applicants submit that support for the language noted by the Office in the rejection is found in the above-identified application at page 4, line 28 through page 5, line 4; page 21, lines 5-18; and claims 8 and 12 as originally filed. In view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw this rejection.

The Office has rejected claims 1-10, 35 and 36 under 35 U.S.C. §112, second paragraph, as being indefinite. The Office asserts that the terms “relative frequency,” “tuple” and “workload” found in the language, “wherein the usage of a given *tuple* relates to a *relative frequency* with which the given tuple was accessed by the queries in the *workload*,”

- 9 -

as recited in claims 1, 7, 10 and 35, are indefinite because it is not clear how the noted terms are acquired or how they interact with each other (emphasis added). In response, Applicants have amended claims 1-10, 35 and 36 as shown above. Support for these amendments is found in the above-identified application at pages 4-5; pages 19-25; FIGS. 5-6; and claims 1-41 as originally filed. Accordingly, no new matter has been added. Further, the amendments to the claims do not change the scope of the claims and make explicit what was implicit in the prior language by clarifying how the noted terms are acquired and how they interact with each other. In view of the foregoing amendment and remarks, the Office is respectfully requested to reconsider and withdraw this rejection.

The Office has rejected claims 1, 2, 6, 7 and 9 under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 6,026,391 to Osborn et al. (“Osborn”) and claims 3, 4 and 36 under 35 U.S.C. § 103(a) as being unpatentable over Osborn in view of U.S. Patent No. 6,519,604 to Acharya et al. (“Acharya”). The Office asserts that Osborn discloses a method and medium for estimating results of a database query comprising (title): collecting workload information related to queries that have been executed on the database (col. 6, lines 44-50); tracing query patterns of queries in the workload to identify the usage of tuples in the database during execution of the queries (col. 6, lines 23-26); determining sample weights based on tuple usage for each tuple (col. 6, lines 26-30); performing a weighted sampling of the database based upon the sample weights (col. 7, lines 17-22); and executing the database query on the weighted sample to estimate results of the database query (col. 7, lines 17-22). Applicants submit the following remarks in response to these rejections.

Neither Osborn nor Acharya, alone or in combination, suggest or disclose, “tracing query patterns … to identify a relative frequency with which at least one given tuple … was accessed in executing the queries …,” as recited in claims 1, 7 and 35, or “a module that traces query patterns … to identify a relative frequency with which at least one given tuple with respect to at least one other given tuple was accessed in executing the queries …,” as recited in claim 10. In general, Osborn discloses estimating a CPU time for executing a present query based on the actual CPU times for prior executed queries that are similar to the present query. The Office’s attention is respectfully directed to Osborn at FIG. 3 and col. 6, lines 51-64, which illustrate and describe a query history 50 table with rows of information representing prior executed database queries 58. The information contained in the columns

- 10 -

of each row of the query history 50 table includes, *inter alia*, a result set 66 and an estimated cost 70 for executing each query 58. Referring now to FIG. 4 and col. 7, lines 7-10 in Osborn, the query history 50 table shown in FIG. 3 is evaluated to find a query 58 that matches a present query 40. Specifically, the cost estimate 44 and the result set 45 for the present query 40 are compared to the result set 66 and the estimated cost 70 for each query 58 in the query history 50 table. When a match is found, then the actual CPU time 72 of the matching query 58 is selected as the estimated CPU time 82 for the present query 40, as disclosed at col. 7, lines 10-16. If no match is found, however, then a nearest neighbor algorithm 83 is employed to extrapolate an estimated CPU time 84 for the present query 40 based on the weighted average of CPU times for the closest matching queries 58, as disclosed at col. 7, lines 17-23.

However, Osborn does not suggest or disclose, evaluating the rows (i.e., queries 58) in the query history 50 table to identify the frequency with which the rows were accessed during the execution of queries. As discussed above, just the cost estimate 44 and the result set 45 for a present query 40 are compared to the result set 66 and the estimated cost 70 for each prior executed query 58 in the query history 50. The result set 66 represents an execution plan for a particular query 58 and is based on available access paths to be used for executing a query, as disclosed at col. 6, lines 23-26. The estimated cost 70 represents the relative system resource cost of the execution plan (i.e., result set 66) and is based on the data distribution and storage characteristics for the respective tables, clusters and indexes to be used for executing a query 58, as disclosed at col. 6, lines 26-29. The result set 66 and the estimated cost 70 represent the particulars with respect to how each query 58 in the query history 50 was executed, not the frequency with which the queries 58 were accessed during execution. Like Osborn, Acharya fails to disclose or suggest, “tracing query patterns ... to identify a relative frequency with which at least one given tuple ... was accessed in executing the queries ...,” as claimed.

As discussed in the above-identified application at page 4, lines 9-11, the low selectivity of queries can contribute to significant error in approximating aggregate values since no single sample of the database can answer all queries with low selectivity with sufficient accuracy. If the selectivity is low, then it dramatically and adversely impacts the accuracy of sampling-based estimation, as stated at page 4, lines 12-13 in this application.

When few relevant tuples are used for estimating a result for a query, significant error is introduced, as set forth at page 4, lines 17-18. Referring now to page 4, lines 28-31 in the above-identified application, workload information associated with prior executed queries is obtained and analyzed to determine which tuples are accessed more often in the workload. This allows the tuples to be used for estimating the query result to be tuned by weighting the tuples based how often they were accessed in the workload during prior execution of queries.

The weighted tuples that are accessed more often are sampled at a higher rate for estimating a query result than the weighted tuples that are accessed less often. Using this weighted sampling approach causes more relevant tuples to be used for query result estimation, and fewer sampled tuples to be discarded due to not meeting selection criteria where the access patterns of the workload are local, and the workload is a good representation of the actual queries that will be posed in the future, as set forth at page 4, line 33 through page 5, line 4. By tuning the sample used for estimating the query result based on a representative workload (i.e., set of queries), the queries can be estimated more effectively, as set forth at 19, lines 20-22 in the application. In view of the foregoing remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 1, 7, 10 and 35. Since claims 2-6 depend from and contain the limitations of claim 1, claims 8-9 depend from and contain the limitations of claim 7, and claim 36 depends from and contains the limitations of claim 35, they are patentable in the same manner as claims 1, 7 and 35.

Additionally, new claims 42-48 are distinguishable from and are patentable over the cited references for the following reasons. Neither Osborn nor Acharya, alone or in combination, disclose or suggest, "obtaining a usage frequency for at least one value within the set of values based on how often the at least one value was used to calculate a result in at least one executed query," as recited in claim 42. As discussed above, Osborn does not teach evaluating the rows (i.e., queries 58) in the query history 50 table to identify the frequency with which the rows were accessed during the execution of queries since just the cost estimate 44 and the result set 45 for a present query 40 are compared to the result set 66 and the estimated cost 70 for each prior executed query 58 in the query history 50. Like Osborn, Acharya fails to disclose or suggest, "obtaining a usage frequency for at least one value within the set of values based on how often the at least one value was used to calculate a result in at least one executed query," as claimed.

- 12 -

Again, obtaining the workload information associated with prior executed queries enables tuning the samples to be used for estimating the query result based on the groups that are accessed more often in the workload than other groups, as set forth at page 4, lines 28-31 in the above-identified application. By tuning the sample used for estimating the query result based on a representative workload (i.e., set of queries), the queries can be estimated more effectively, as set forth at 19, lines 20-22 in the application. In view of the foregoing, claim 42 stands in condition for allowance. Since claims 43-48 depend from and contain the limitations of claim 42, they are patentable in the same manner as claim 42.

In view of all the foregoing, it is submitted that this case is in condition for allowance and such allowance is earnestly solicited. In the event that there are any outstanding matters remaining in the above-identified application, the Office is invited to contact the undersigned to discuss this application.

Respectfully submitted,

MICROSOFT CORPORATION

Date: June 30, 2004

By *John Campa*  
John Campa  
Reg. No. 49,014  
Direct Tel. No. (425) 706-0731

**Attachment**